Running head: Expressive and Receptive Language Effects

Expressive and Receptive Language Effects of African American English on a Sentence Imitation Task

J. Michael Terry, Ph.D., Assistant Professor

Department of Linguistics

University of North Carolina Chapel Hill, Chapel Hill, NC

Sandra C. Jackson, Ph.D., CCC-SLP, Associate Professor

Department of Communication Disorders

North Carolina Central University, Durham, North Carolina

Evangelos Evangelou,Ph.D., Lecturer

Department of Mathematical Sciences

University of Bath, Bath, UK

Ricahrd L. Smith, Ph.D., Professor

Department of Statistics and Operations Research

University of North Carolina Chapel Hill, Chapel Hill, NC

Corresponding Author:

J. Michael Terry
Department of Linguistics
Campus Box 3155, UNC Chapel Hill, NC 27599-3155
Chapel Hill, NC 27599-3155
jmterry@unc.edu

Abstract

This study tests the extent to which giving credit for African American English (AAE) responses on a General American English (GAE) sentence imitation test mitigates dialect effects.

Forty-eight AAE-speaking second graders completed the Recalling Sentences subtest of the Clinical Evaluation of Language Fundamentals-3 (CELF-3, 1995). A Bayesian Markov Chain Monte Carlo method was used to determine the relationship between the students' scores and the presence of third person singular -s, a feature largely absent from AAE morphosyntax, in the subtest sentences.

Even when given credit for AAE responses, the estimated effect of third person singular -s was significant, high relative to those of negation and counterfactual conditional *if+ed*, and correlated with an independent measure of the students' rootedness in AAE syntax.

It is argued that these results reveal a receptive language effect not addressed by crediting dialect productions.

KEY WORDS: African American English, Assessment, Dialect, Morphosyntax, Sentence Imitation.

Expressive and Receptive Language Effects of African American English on a Sentence Imitation Task

The primary goal of this paper is to analyze the extent to which giving speakers of African American English (AAE) credit for alternate AAE responses to items on the Recalling Sentences subtest of the Clinical Evaluation of Language Fundamentals-3 (CELF-3, 1995) mitigates dialect effects. One of six core subtests of the CELF-3, Recalling Sentences asks students to imitate General American English (GAE) sentences read aloud to them by the examiner. Although the CELF-4 (Semel, Wiig, & Secord, 2003) is a more current version of the CELF, and increases the Recalling Sentences subtest from 26 to 32 items, the task performed on the CELF-3 is the same as that of the CELF-4. Further, the 4-point scoring scale used on the CELF-4 (3, 2, 1, 0) is inherited and remains unchanged from the CELF-3. Consistent with the CELF-3 manual's directive to "count dialectal variation as correct" on any expressive language subtest "if the variation is appropriate for that student", the list of alternate AAE responses used in this study is the same as that suggested in the CELF-3 manual except that additional features that are consistent with AAE morphosyntax are included (Wolfram, Jackson & Roberts, 2000).

This list (given in Appendix 1) focuses on morphosyntactic variations. Mounting evidence suggests not only that AAE affects performance on a variety of tests given in GAE, but that effects largely track morphosyntactic differences between the two dialects (Terry, Hendrick, Evangelou, and Smith in press; Craig & Washington, 2006). Typically, the impact of such differences is viewed as primarily an expressive language issue, and where testing is concerned, accounted for solely in that domain. Note for instance that the CELF-3 does not make allowances for dialect on any of its receptive language subtests, only those that test expressive language. In similar fashion, the AAE allowances investigated in this paper exclusively address expressive language difficulties. We see the use of these allowances, then, as representative of a general, and not uncommon, strategy for accounting for dialectal differences in similar testing situations. Our analysis of this strategy is likewise meant to reveal something general. The key question we are concerned with is whether morphosyntactic mismatches between GAE and AAE have significant receptive language effects for AAE speaking students on GAE sentence imitation tests that leave this strategy wanting. Receptive language is meant here in its broadest sense, and not simply as a synonym for semantic comprehension. As a result, it includes the encoding of structural features of language that, depending on one's take, may or may not be considered part of meaning. Our principal question, then, is whether morphosyntactic mismatches between GAE and AAE interfere with AAE speaking students' ability to encode the information and or structures needed to perform GAE sentence imitation tasks. Recent work by Terry, Hendrick, Evangelou, and Smith (in press) presents evidence that morphosyntactic mismatches between GAE and AAE do have significant effects on second grade AAE speaking students' performance on mathematical reasoning tasks. That work follows from the conjecture that bilingual and bi-dialectal speakers, to the extent that they are required to switch between linguistic codes in verbally mediated tasks, have an added cognitive load that can have variable and measurable effects. The effects reported are taken to support that view (which should in no way be confused with the position that AAE lacks the ability to capture complex mathematical concepts.) Whether similar effects can be found on sentence imitation tests is an open question.

Using the same statistical machinery and drawing from the same population of students, in the current study, we wish to do three things. First, we offer support for the view that morphosyntactic differences between AAE and GAE have measurable effects on sentence imitation tests such as the CELF-3 Recalling

Sentences subtest. Second, we show that allowances for dialectal difference, such as advocated in the CELF-3 manual and given in the Wolfram, Jackson & Roberts (2000) list of alternatives used here, do help mitigate these effects. Finally, we argue that while these allowances make up for some of the effect of feature mismatches, they do not make up all of it. We further argue that this is because these allowances only account for expressive language effects, whereas the effects of morphosyntactic feature mismatches are both expressive and receptive.

Method

Participants

Participants were 48 second grade students (24 boys/25 girls) drawn from a pool of 87 African American children originally recruited from North Carolina community based child care centers to participate in a longitudinal study of children's health and development (Roberts, et al., 1995). The children were speakers of AAE (Jackson & Roberts, 2001) and had no known biological or genetic risks. As a part of the original study, at regular intervals, language samples were taken from the children, and they were administered a series of diagnostic tests to assess their linguistic and other cognitive abilities.

Of central importance to this study, these tests included the Recalling Sentences subtest of the Clinical Evaluation of Language Fundamentals 3 (CELF-3). This subtest along with other age-appropriate subtests of the CELF-3, was given at the end of each grade year beginning with the first grade. Data for this study were collected when the children were in the second grade. The mean age of the students at testing was 8.35 (SD = .83)

Procedure

All measures were collected and tests administered by one of seven trained examiners with expertise in speech and language assessment. The CELF-3 and narratives that were used in this study and all other measures were collected in a room at a university research facility.

Sentence imitation task. Designed to differentiate normal from disordered language in both the expressive and receptive language domains, the CELF test battery is used to identify children, adolescents, and young adults who lack the fundamentals of form, content, and (linguistic) memory that distinguish mature language use. Recalling Sentences is a subtest that evaluates the student's ability to listen to and repeat sentences of increasing length and complexity without changing their syntactic, semantic or morphological properties. Sentence attributes such as non-canonical word order, negation, and multiple clauses are used to add to the complexity of sentences and make the task more difficult. An inability to perform this task is taken to show that a student has either failed to retain or represent in sufficient strength key meanings or structural features that are necessary for accurate recall and repetition. There are 26 items on the CELF-3 with a 4-point scoring scale (3 = 0 errors, 2 = 1 error, 1 = 2 - 3 errors, and 0 = 4+ errors). Students were given credit for alternate AAE responses as guided by those suggested in the CELF-3 manual except that additional features that are consistent with AAE morphosyntax are included (Wolfram, Jackson & Roberts, 2000). A full list of accepted variations is given in Appendix 1.

Collection of language samples. Narrative language samples were collected and used to establish each student's level of AAE usage. The examiner elicited three narratives from each child as follows: telling a story about bears after viewing slides on a personal slide viewer, describing a picture of a circus, and responding to personal narrative prompts about situations such as telling about having lost a tooth. Each narrative was transcribed using Systematic Analysis of Language Transcripts (SALT) (Miller & Chapman 2000).

Scoring of AAE using language samples. The narrative language samples just described play an important role in the current study and were coded for a set of 44 morphosyntactic AAE features. Dialect Density Measures (DDMs) were calculated from each student's narratives to establish his or her AAE speaking status. A DDM expresses the rate of dialect feature production calculated as a ratio of number of dialect features to number of words (Craig & Washington 2004, Oetting & McDonald 2002). Narratives had to meet a 50 word threshold before a DDM was calculated. The mean DDM score for the students in this study was 0.169 with a standard deviation of 0.121. Although, as these numbers indicate, there was some variation in the scores, all of the students had scores indicating that they were AAE speakers. For comparison purposes, consider that using a hierarchal cluster analysis, Washington, Craig, and Kushmaul (1998) were able to group 65 4;4 - 6;3-year old African American children into level-of-AAE-use categories in two different sampling contexts, free play and during a picture description task. Free play yielded two groups: "high" and "low". Scores for their high use group ranged from .082 to .161 (SD = .017); scores for their low use group ranged from .024 to .073 (SD = .023). The picture naming task yielded more groups. Scores for their "very high" use group ranged from .177 to .196 (SD = .010); the "high" group scores ranged from .094 to .196 (SD = .010); the "moderate" group scores ranged from .050 to .091 (.010); and the "low" group scores ranged from .041 to .031 (SD = .010).

Reliability. Above 90% reliability between coders was established over the full set of subtest questions for the CELF-3 and for the narrative language samples. In addition to data concerning individual students' performance on CELF-3 Recalling Sentences subtest questions, two members of our team coded each of the 26 subtest questions for a range of linguistic properties such as whether or not a question included conditional if+ed structure or negation. The two coders were 100% reliable with one another.

Results and Discussion

Three Morphological Features Chosen for Data Analysis

Three morphological features were chosen for further statistical analysis. The first, third person singular -s, was chosen because it has been identified as a point of divergence between AAE and GAE (Green 2002, Craig and Washington 2006) and because previous work suggests that its presence may affect AAE speakers performance on a variety test types. In Terry, Hendrick, Evangelo & Smith in press, it is argued that third person singular -s has an educationally significant effect on second grade AAE speaking students' performance on the WJ-R Applied problems subtest. Similarly, Labov and Charity-Hudley (2009) find that the presence of third person singular -s in a passage creates online disturbances and misreadings further downline from the feature itself for AAE speaking readers. Not being dialect features, the other two linguistic features, counterfactual conditional *if+ed* and negation, were selected for use in the current study for comparison purposes. Both are assumed to add difficulty to the sentence imitation items. Negation has been shown to pose cognitive difficulties on a variety of tasks (Sherman 1973, Clark 1974), and counter factual *if+ed* introduces multiple clause structure, making the task syntactically more complex. All three features are listed along with GAE examples of their use in Table 1.

[Insert Table 1 Here]

In line with previous research and consistent with the instructions on the CELF-3 to give credit for dialectal responses, morphosyntactic mismatches between GAE and AAE were hypothesized to have a measurable effect on students' performance on the Recalling Sentences subtest. Such feature effects were assumed to be both expressive and receptive in nature, and it was further predicted that the allowances for alternate AAE responses would only partially make up for them. Receptive language effects operate across test sentences, making entire sentences more difficult to repeat, whereas the expressive allowances only operate at the word-level, having an impact only when an alternate form of a word is produced. For example, an inability to encode the meaning or structural features associated with the -s in the sentence *Bobby plays with toy trucks* can affect a test takers ability to repeat the entire sentence. Giving credit for the *Bobby play with toy trucks* as a repetition of the original sentence only has an effect if the word *play* is substituted for *plays*.

To test the initial hypothesis that the feature mismatches would have any kind of effect, we sought to determine whether the performance of the AAE speaking children on the Recalling Sentences subtest correlated with third person singular -s, the first of the linguistic features in Table 1.

An MCMC Model of Linguistic Effects

As noted previously, the CELF-3 recognizes four scoring levels on each of the Recalling Sentences items. Students are awarded a score of three for each sentence repeated without error, two if the repeated sentence contains one error, one if the sentence contains two to three errors, and zero if it contains four or more errors. To simplify the statistical model used to analyze performance, two scoring levels were created to account for children's level of success on each of the items. Zero to one error was counted as "high," and two or more errors as "low".

Given a sentence to repeat, we model whether a student receives a high or low score according to the criteria above as a function of the child's recall ability, the difficulty (in terms of recall) of the sentence itself, and the presence or non-presence of any of the linguistic features in Table 1. Recall ability and sentence difficulty are estimated from the data themselves. As we assume that any linguistic feature effects will be cumulative, the model sums the individual instances of each feature. A Bayesian Markov Chain Monte Carlo Method (MCMC) was employed to estimate the unknown parameters, including the effect of the features on a student. Details of the model, including a measure of its fit to the data, are given in Appendix 2. The chief benefit of using the MCMC method is that is allows us to treat the model parameters (the student's general recall ability, the overall difficulty of the question, and the extent to which the student is affected by a given linguistic feature) as random effects. If one treats the influence of the linguistic features as fixed rather than random effects, (i.e. if one assumes the influence of a particular linguistic feature is the same for all students), it is possible to estimate the model by standard logistic regression. However, not only is the number of unknown parameters too large to assume fixed effects, but doing so prevents testing whether a student's rootedness in AAE syntax correlates with the effect that the AAE features have on test performance, an important hypothesis pursued here.

Scoring Without Allowances for AAE Responses

In determining whether the features in Table 1 influence students' performance, we first consider scoring in which no allowances are made for AAE responses. For each linguistic feature we examined, Table 2 provides α_{ik}^{M} - the estimate of the influence of that feature on the student to perform at a high or low level averaged across students - along with its standard deviation.

[Insert Table 2 Here]

We see in Table 2 that of all the features examined, third person singular -s has the α_{ik}^{M} with the greatest absolute value and thus appears to have the greatest effect on students as a whole. In comparison, counterfactual conditional *if+ed*, which introduces syntactic complexity by introducing multiple clause structure, and negation, which introduces semantic complexity, have lower values of α_{ik}^{M}.

An alternate way to see the effect of the features on student performance is to consider the correlation between students' overall scores and α_{ik}^{M}, the influence of a feature on the students as a whole. For each linguistic feature we examined, Table 3 provides the correlation between a students' total score on the CELF-3 Recalling Sentences subtest (calculated using the success levels previously described) and α_{ik}^{M}, the measure of the influence of that feature on the student provided by the model. If there is no correlation between a student's score and the linguistic feature, we expect the correlation to be 0. A high positive or negative value indicates that the effect of the feature is high. In the case of a positive correlation, students who are strongly negatively affected by the linguistic feature in question have worse than average scores, while students who show a high positive effect, tend to have strong scores. The reverse is true in the case of a negative correlation. Table 2 also provides the p-value for the null hypothesis that there is no correlation between a student's score and the linguistic feature; the lower and upper bound for the 95% confidence interval for the correlation is also reported.

[Insert Table 3 Here]

Table 3 shows a relatively strong correlation between the effect of third person singular -s on students and their overall scores on the Recalling Sentences subtest. Compared to the counterfactual *if+ed* construction and negation, which we expect to affect scores, the correlation for third person singular -s is higher. Taken together, the data in Table 2 and Table 3 constitute evidence that the presence of third person singular -s has a significant effect student performance.

Scoring with Allowances for AAE

We now consider what effect the AAE allowances have on student scores and our estimates of linguistic feature effects. Similar to Table 2, for each linguistic feature, Table 4 provides α_{ik}^{M} - the model estimate of the influence of that feature on a student to score high or low, averaged across students - along with its standard deviation. Again, high level performance is defined as making zero to one changes to the sentence being repeated; low level performance is making two or more changes. These numbers, however, are for the case in which students are not penalized for substituting any of the AAE morphosyntactic features listed in Appendix 1 for their GAE equivalents.

[Insert Table 4 Here]

Table 4 shows a sharp decline in the estimated influence of both third person singular -s and counterfactual *if+ed* on the students in this study. As nothing else changed from Table 2 to Table 4, in both cases, this decline can be attributed to the AAE allowances. Were there no AAE feature variations being penalized in the first accounting, there would be no change in the second. In the case of third person singular -s, the most likely explanation is that the variation accounted for was due to the absence of third person singular -s in students' sentence repetitions. In the case of sentences containing *if+ed* structures, since at least one of the forty eight students received an AAE allowance on each of the sentences that contained an *if+ed* structure, it is possible that its influence was overestimated in Table 2. In contrast, there was only one instance of a student receiving more than one allotment for any for the four sentences that contain third person singular -s.

While α_{ik}^{M} values for both third person singular -s and *if+ed* declined sharply, the α_{ik}^{M} for -s remains high relative to that for *if+ed*. This might be explained in at least two ways. First, it is possible that there is some sentence attribute that is strongly correlated with the presence of third person singular -s and not accounted for in the general problem difficulty term of the modeling equation that is influencing students. Second, it could be that some aspect of the influence of third person singular -s, presumably a receptive language effect, is not being accounted for by the scoring allowances. That the two things are happening at once is also a possibility. We give further attention to the interpretation of these numbers shortly.

We turn now to the correlation between α_{ik}^{M} and students' overall success/failure score. Similar to Table 3, Table 5 provides this correlation along with the p-value for the null hypothesis that there is no correlation between a student's score and the linguistic feature, and the lower and upper bounds for the correlations 95% confidence interval.

Whereas Table 4 showed a decrease in the strength of the α_{ik}^{M}s from the scoring without AAE allowances case (especially for that of third person singular -s), in Table 5, we see an increase in the strength of the correlations with score. This pattern is to be expected if expressive language effects have been accounted for, but receptive language effects have not. The overall score reflects both word-level and sentence-level effects since repetition errors are the result of both effect types. It is likely, however, that these two effect types combine in a non-linear fashion. That is, a student is more likely to make a single error that results from a combination of word and sentence-level effects reaching some threshold than one error due to word-level effects and one due to sentence-level effects. Thus, adjusting for expressive errors (the result of word-level effects) has the result of making the data more linear and increasing the correlation with score.

[Insert Table 5]

The data in Table 5 are consistent with the hypothesis that there is a residual sentence-level effect of third person singular -s that is not accounted for by the use of alternate

AAE allowances.

A Residual Effect

Thus far we have presented evidence that third person singular -s has a significant effect on AAE speaking second graders' scores on the CELF-3 Recalling Sentences subtest. Much of this effect is mitigated by the use of a list of alternate AAE responses, which gives students credit for producing sentences that, consistent with AAE grammar, show absence of this linguistic feature. Using this list of alternate AAE responses, however, does not eliminate the effect. In fact, the estimated effect remains high compared to those of negation and counterfactual *if+ed,* two linguistic features assumed to make sentences harder to recall. One possible explanation for this "left over" effect is that morphosyntactic mismatches between AAE and GAE have two types of effects on sentence recall tests: word level effects and broader sentence-level effects. Word-level effects reveal themselves when students use (or do not use) particular morphemes. This is the type of effect that can be accounted for by using a list of alternate AAE responses as they show up in expressive language measures. Sentence-level effects, on the other hand, are more likely to be receptive language effects. Here, morphosyntactic mismatches add to the overall difficulty of the repetition task by introducing an extra cognitive load. Although this explanation is the driving hypothesis of the current paper, the data in the previous section are only consistent with it. To look for positive evidence for this view we introduced an independent measure of the students' entrenchment in the syntax of AAE grammar into our modeling equation. Given a sentence on the Recalling Sentences subtest, we treat the level of a student's success, then, as a function of this measure, the student's recall ability, the recall difficulty of the sentence, and the presence or non-presence of the linguistic features under consideration. The idea here is that if the residual effect of third person singular -s is related to dialect, then this new term should account for some of the variance in a high or low level of success, and thus, lower the estimated effect of the feature.

For this new measure we used students' rates of copula omission calculated from their unscripted narratives. Dialect density measures such as was used initially to determine the students' AAE status are too broad, containing both phonological features that may not say anything about a student's rootedness in AAE syntax and morphosyntactic features that may be structurally dependent on third person singular -s. Under the Labovian deletion analysis, copula absence, while a quintessential feature of AAE syntax, is not structurally related to third person singular -s absence (Labov 1969).

Table 6 shows α_{ik}^{M}, the estimated linguistic feature influence on student success averaged over students along with its standard deviation under the model assumptions just outlined.

[Insert Table 6]

Table 7 gives the correlation between α_{ik}^{M} and overall score along with the p-value for the null hypothesis that there is no correlation between a student's score and the linguistic feature, and the lower and upper bounds for the correlations 95% confidence interval.

[Insert Table 7]

Table 6 shows the predicted reduction in the mean α_{ik}^{M} for third person singular -s. Table 7 shows that its correlation with score is still significant. Taken together this suggests that there is a significant effect of third -s, related to dialect, but not the production of third -s itself. This is expected if there are significant receptive language effects of third -s, not accounted for in by the AAE allowances. Less expected and less easy to explain in Table 6 is the reduction in the mean α_{ik}^{M} for counter factual *if+ed.* One possible explanation is suggested by findings noted in Craig and Washington (2004). They report that dialect switching between AAE and mainstream English is typically accompanied by reduced sentence complexity on the part of the speaker. This suggests to us that dialect switching is purchased at the cost of less linguistic complexity, while linguistic complexity is bought with less dialect switching. As we have used the *if+ed* structure as a measure of complex syntax, it may be that the effect of *if+ed* is particularly high for students deeply rooted in AAE syntax because they must make this trade off, and the introduction of a measure of their rootedness into the model allows this effect to be redistributed.

Another issue raised by the results in Table 6 is whether a significant effect of third person singular –s remains after introducing the new model. It might, for instance, be the case that third person singular -s presents a cognitive load for AAE speakers and non-AAE speakers alike, with AAE speakers being particularly affected due to the structure of the dialect. Under this account one would expect a residual effect of the feature to appear in Table 6 reflecting a load unrelated to dialect. Unfortunately, because mean measure of effect and its standard deviation are so close in absolute value, it is impossible to tell with any certainty whether there is still a significant effect of this feature. The same is true of *if+ed.* Further work is required to determine what if any effects of these features might remain.

This paper presents evidence that morphosyntactic differences between AAE and GAE have measurable receptive as well as expressive language effects on tests of sentence recall such as the CELF-3 Recalling Sentences subtest. In doing so, it draws attention to the more general issue of receptive language effects of dialectal difference. An interesting connection might be, for instance, be made to Restrepo and colleagues (2006) findings that four year old African American children tend to perform better on expressive tests of vocabulary than on receptive ones. In that work, children's scores were worse on tests that asked them to demonstrate knowledge of an examiner-provided word by selecting one of four pictures than on tests in which they themselves were required to produce a word.

Currently, allowances for dialectal difference such as advocated in the CELF-3 manual, and given in the alternate response allowances examined here, only account for expressive language effects on sentence imitation and numerous other types of tests. Further research is needed to better quantify receptive language effects and to seek adequate remedies for the diagnostic difficulties that they pose. Careful attention to these issues, and likely more careful attention to the wording of language tests and reading assessments are needed. No test, however, can replace the SLP. Test results must be interpreted and their interpretations used to develop viable academic and therapeutic plans. While more research is being done on the nature of these receptive effects, practicing SLPs should be aware that they exist so they can better interpret the results of the tests they use, and better distinguish between dialectal difference and language disorder in the linguistically diverse communities in which they provide speech and language services.

REFERENCES

Clark, H. (1974). Semantics and Comprehension. In T.A. Sebeok (Ed.), Central Trends in Linguistics, vol 12. Linguistics and Adjacent Arts and Sciences. The Hague: Mouton.

Clinical Evaluation of Language Fundamentals-Third Edition. (1995). San Antonio: The Psychological Corporation, Harcourt Brace & Company.

Craig, H. K., & Washington, J. A. (2004). Grade-related changes in the production of African American English. *Journal of Speech, Language and Hearing Research, 47,* 450-463.

Craig, H. K., & Washington, J. A. (2006). *Malik goes to school: Examining the language skills of African American Students from preschool --5th grade*. Mahwah, NJ: Lawrence Erlbaum.

Jackson, S. C. & Roberts, J. E. (2000). Complex syntax production of African American Preschoolers. Journal of Speech, Language, and Hearing Research, 44, 1083-1096.

Johnson, V. E., Devilliers, J. G., & Seymour, H. N. (2005). Agreement without understanding? The case of third person singular /s/. *First Language.*

Labov, W. (1969). Contraction, deletion and inherent variablity of the English copula. *Language, 45,* 715-762.

Labov, W., & Charity-Hudley, A.H. (2009). Symbolic and structural effects of dialects and immigrant minority languages in explaining achievement gaps. Paper presented at the National Academy Division on Behavioral and Social Sciences and Education Workshop on the Role of Language in School Learning: Implications for Closing the Achievement gap, Menlo Park, CA.

Miller, J., Chapman, R (2000). Systemic Analysis of Language Transcripts (SALT) [Computer Software]. University of Wisconsin,Language Analysis Lab, Madison, WI.

Oetting, J. B. and McDonald J.(2002). Methods for characterizing participant's non-mainstream dialect use within studies of child language. *Journal of Speech, Language, and Hearing Research,* 45, 505-518.

Restrepo, M. A., Schwanenflugel, P. J., Blake, J. Neuharth-Pritchett, S., Cramer, S.E., & Ruston, H. P. (2006). Performance on the PPVT-III and the EVT: Applicability of the measures with African American and European American preschool children. *Language, Speech, and Hearing Services in Schools*, *37*, 17-27.

Roberts, J. E., Burchinal, M., Medley, L. P., Zeisel, S. A., Mundy, M., Rousch, J., Hooper, S., Bryant, D, & Henderson, F.W. (1995). Otis media, hearing sensitivity, and maternal responsiveness in relation to language during infancy. *Journal of Pediatrics, 126*(3), 481-489.

Semel, E., Wiig, E. H., & Secord, W. (2003). *Clinical evaluation of language fundamentals-4*. San Antonio: PsychCorp,Harcourt Assessment, Inc.

Sherman, M. (1973). Bound to be easier? The negative prefix and sentence comprehension. Journal of Verbal Learning and Verbal Behavior, 12, 76-94.

Terry, J.M., Hendrick, R., Evangelou E, & Smith R.L. (in-press). Variable dialect switching among African American children: inferences about working memory. *Lingua*.

Torrey, J. (1972) The language of Black children in the early grades. New London: Connecticut College.

Washington, J. A., & Craig, H. K. (2001). Reading performance and dialectal variation. In J.L. Harris, A.G Kamhi & K.E. Pollock (Eds.), *Literacy in African American communities* (pp. 147-168). Mahwah, NJ: Erlbaum.

Washington, J. A, Craig, H., & Kushmaul, A. (1998). Variable use of African American English Across two language sampling contexts. *Journal of Speech, Language, and Hearing Research,* 41, 1115-1124.

Waters, G. S., & Caplan, D. (1999). *The effect of a digit-load on syntactic processing in subjects with high and low working memory capacity.* Unpublished manuscript.

Wolfram, W., Jackson, S.C., & Roberts, J. E. (2000). Clinical Evaluation of Language Fundamentals-3 alternate African American English Responses.

Young, G. A., & Smith, R. L. (2005). *Essentials of statistical inference.* Cambridge: Cambridge University Press.

APPENDIX A

The following are examples of the AAE allotments given in the study. The targets in the study were taken directly from the CELF-3 Recalling Sentences subtest. Those given here are only modeled on them. Full credit was given for each item (score of 3), except where errors are noted in parentheses

Target	**Acceptable Responses**
The car was followed by the ambulance.	The car was **follow** by the ambulance The tractor **( )** was followed by the ambulance (1 error)
The baby was not put in the crib.	The baby **won't** put in the crib The baby **( )** not put in the crib
Was the car preceded by the truck?	Was the car **precede** by the truck?
The short eighth grader made the basket.	The short eighth grader **make** the basket.
The man caught the ball and the woman booed loudly.	The man **catch/catched** the ball and the woman booed **loudly/loud.**
The girl stopped to buy some milk, even though she was late for school	The boy **stop** to buy some milk … … **( )** though she was late for school.
The girl loaned a book to her friend who like short stories.	The girl loaded a book to her friend **( )** likes/**like** short stories/story. … **that** likes/like short stories/story
The boy did not know the teacher who taught fifth grade last year.	The boy did not/**ain't** know the teacher **( )** taught fifth grade last year. …**that** taught/**teach** fifth grade …
The tables and chairs were donated by the school board.	The table(s) and chair(s) **was** donated by the school board.

The boy who lives downstairs in my apartment building is on my team.	The boy **( ) lives**/live downstairs … The boy **that** lives/live downstairs .. … in my apartment building **( )** on my team.
If the man had baked some brownies, they would have been eaten by now.	If the man **( )** baked some brownies … If the man **bake** some brownies If the man has bake … (1 error) … they would **( )** be eaten by now. … they would **( )** be ate by now. … they would **( )** been eaten by now. … they would **( )** be aten … (1 error) ..they would **( ) ( )**eaten … (1 error) … they **( )** have been eaten … (1 error) … they **( )**be **eat/et** ... (2-3 errors).
The girl who won the award at band camp was very happy.	The girl **( )** won the award at band camp **she** was very happy The girl **that** won the award …
If the snow doesn't stop before eight, school we have to be cancelled.	If the snow **don't** stop before, eight, … school **is gonna** have to be cancelled. …school **() () () be** cancelled (1 error) … school **would () ()** be cancelled. (2 errors)
The boys mailed a package to their friend who moved away last year.	The students mail a package to they friend that moved/**move** away last year. … friend ( ) moved/**move** …

The man who helps collect the lunch money at school is the new librarian.	The man **that** helps/**help** collect … The man **( )** helps/help collect … … at school **( )** the new librarian.
The men collected and repaired the TVs, and sold them in the store.	The men **collect** and **repair** the TVs, and then **sold** them in the shop. … **selled** them in the shop.
After the children had completed the lesson the teacher asked them to write a paper.	After the students **complete** the lesson, the teacher **ask** them to write a paper.
Before the juniors were dismissed for the day, they were told to turn in their reports.	Before the juniors were/**was dismiss** for the day, they were/**was** told to turn in their/**they** reports.
The girl who didn't show up for practice wasn't allowed to play volley ball until two weeks later.	The girl that didn't'/**ain't** show up for practice … The girl **( )** didn't'/**ain't** show up for practice … … wasn't/**won't** allowed to play on the team until/till two weeks later.
If the teacher had let us top class earlier, we would have been home long ago.	If the teacher **( )** let us stop class … … we would **( )** been home long ago. … we **( ) ( )** been home … (1 error) … we **( ) ( )** be home … (1 error)
When the children had finished playing, they decided to get something to eat before going home.	When the children had **finish** playing … When the children **( )** finish/**finish** studying … … they **decide** to get something to eat before going home.

The English teacher sorted, labeled, boxed, and delivered the books.	The English teacher **sort**, **label**, **box** and deliver the books.
The teacher in the room next door promised to water the plants during our summer vacation.	The teacher in the room next door **promise** to water the plants during our summer vacation.

APPENDIX B

In this appendix we outline a Bayesian-MCMC model for the effect of linguistic features on test performance. We treat the model parameters (e.g., the overall difficulty of the question, the extent to which the student is affected by a linguistic feature, etc) as random effects. The Bayesian-MCMC approach to fitting the model is to define prior distributions for these parameters and use (Gibbs and Metropolis) sampling to construct posterior distribution for all the unknowns. The posterior distributions are then used to determine which, if any, linguistic features influence students' scores. By treating the influence of the linguistic features as fixed rather than random effects, (i.e. if one assumes the influence of a particular linguistic feature is the same for all students), it would be possible to estimate the model by standard logistic regression. However, the number of unknown parameters is too large to assume fixed effects. More importantly, to assume that they were fixed effects would prevent us from exploring whether the amount of AAE a student uses correlates with the effect that the AAE features have on test performance, an important hypothesis pursued here.

For each student i and sentence j we represent whether the student repeated the sentence at the high success level or the low success level with z_{ij} ; z_{ij} takes the value 1 when student i responds at the high level and 0 when the student responds at the low level. We treat any skipped sentences as missing values and ignore them. We define y_{ij} to be the command of student i of sentence j. The variable y_{ij} is an unobserved random variable such that $z_{ij} = 1$ if $y_{ij} > 0$ and $z_{ij} = 0$ if $y_{ij} \leq 0$. As we are interested in the effect of the linguistic features on students' responses, we define α_{ik} as the influence of the linguistic feature k on student i. Because each linguistic feature can potentially appear multiple times in a given test question, we let x_{jk} represent the number of times that linguistic feature k appears in sentence j Other factors besides the presence or non-presence of the linguistic features will influence students' responses to the test questions. For example, we assume that a student's overall recall ability plays a role, as does the difficulty of the question. We represent these influences as η_i and β_j respectively. We use the model in (1), for $i = 1,\ldots,48$ and $j = 1,\ldots, 28$.

(1) $y_{ij} = \eta_i + \beta_j \sum_{k=1}^{3} \alpha_{ik} \, x_{jk} \quad + \epsilon_{ij}$

The term ϵ_{ij} in the formula represents the error in the model, which we assume is independent of the other variables and is produced by factors we have not taken into account.

We assume that a student's ability, the difficulty of the question, the effect of the linguistic feature on a student's performance and the error of the model are all random variables distributed normally and independently of one another. That is, we assume (2), where μ_η, μ_β, υ_k , σ^2_η , σ^2_β, σ^2_ϵ, τ^2_k , are unknown and need to be estimated from the data along with η_i, β_j, α_{ik} .

(2)

$\eta_i \sim N(\mu_\eta, \sigma^2_\eta)$, $i = 1,\ldots,48$

$\beta_j \sim N(\mu_\beta, \sigma^2_\beta)$, $j = 1,\ldots,26$

$\epsilon_{ij} \sim N(0, \sigma^2_\epsilon,)$ $i = 1,\ldots,48$, $j = 1, \ldots,26$

$\alpha_{ik} \sim N(\upsilon_k, \tau^2_k,)$ $i = 1,\ldots,48$, $k = 1,2,3$

To estimate the unknown parameters in the model (2), we apply a Bayesian MCMC method (Young & Smith 2005:22-48). We begin by simplifying the model. Since the model is only affected by the difference between μ_η and μ_β rather than their individual values, we set $\mu_\beta = 0$. Further we set $\sigma_\varepsilon^2 = 1$ since multiplication of all y_{ij} by the same positive constant leaves z_{ij} unchanged. Writing the variance as κ_η^{-1} , κ_β^{-1} , and λ_κ^{-1}(in place of σ_η^2 , σ_β^2 and τ_κ^2)to simplify calculations, the model has the form in (3).[1]

(3)

$$\eta_i \sim N(\mu, \kappa_\eta^{-1}),$$

$$\beta_j \sim N(0, \kappa_\beta^{-1}),$$

$$\epsilon_{ij} \sim N(0, 1),$$

$$\alpha_{ik} \sim N(v_k, \lambda_k^{-1},) \ ,$$

$$y_{ij} = \eta_i + \beta_j + \textstyle\sum_{k=1}^{3} \alpha_{ij}\ x_{jk} + \epsilon_{ij},$$

$$z_{ij} = 1 \text{ if } y_{ij} > 0,$$

$$z_{ij} = 0 \text{ if } y_{ij} \leq 0.$$

The parameters μ, ν_κ, κ_β, κ_η, λ_κ are termed hyperparameters. We set the hyperparameters as in (4) where *U* denotes the uniform

[1] We can do this because a *Gamma* distribution for the prior of κ_η, κ_β and λ_κ gives a *Gamma* posterior distribution for those variables, yielding a distribution in closed form.

distribution and *Gamma* the gamma distribution and $a = b = .001$.[2]

(4)

$\mu \sim U(-\infty, \infty)$,

$\upsilon_{\kappa} \sim U(-\infty, \infty)$,

$\kappa_{\eta} \sim Gamma(a, b)$,

$\kappa_{\beta} \sim Gamma(a, b)$,

$\lambda_{\kappa} \sim Gamma(a, b)$

We are now in a position to specify our MCMC algorithm. Let I represent the number of students (I =48); J is the number of questions (J = 28), and K is the number of linguistic factors (K = 3). We can now treat the joint density of (μ, ν_{κ}, κ_{β}, κ_{η}, λ_{κ}, η_i, β_j, α_{ik}, y_{ij}, z_{ij}) as proportional to (5).

(5)

$$\kappa_{\eta}^{a-1} e^{-b\kappa_{\eta}} \cdot \kappa_{\beta}^{a-1} e^{-b\kappa_{\beta}} \cdot \prod_{k=1}^{K} \left\{ \lambda_k^{a-1} e^{-b\lambda_k} \right\} \cdot \prod_{i=1}^{I} \kappa_{\eta}^{\frac{1}{2}} e^{-\frac{1}{2}\kappa_{\eta}(\eta_i - \mu)^2} \cdot \prod_{j=1}^{J} \kappa_{\beta}^{\frac{1}{2}} e^{-\frac{1}{2}\kappa_{\beta}\beta_j^2} \cdot$$
$$\prod_{i=1}^{I} \prod_{k=1}^{K} \lambda_k^{\frac{1}{2}} e^{-\frac{1}{2}\lambda_k(\alpha_{ik} - \nu_k)^2} \cdot \prod_{i=1}^{I} \prod_{j=1}^{J} e^{-\frac{1}{2}(y_{ij} - \eta_i - \beta_j - \sum_{k=1}^{K} \alpha_{ik} x_{jk})^2} \cdot Q(y_{ij}, z_{ij})$$

where

$$Q(y, z) = \begin{cases} 1 & \text{if } y > 0 \text{ and } z = 1, \\ 1 & \text{if } y \le 0 \text{ and } z = 0, \\ 0 & \text{otherwise.} \end{cases}$$

With the exception of z_{ij} , all the variables in (5) are

---

[2] We assume that these hyperparameters are independent of each other and of the effects of α, η, and β.

unknown and need to be estimated. The Bayesian approach to this problem is to construct the conditional density of (μ, ν_κ, κ_β, κ_η, λ_κ, η_i, β_j, α_{ik}, y_{ij}, z_{ij}) given all z_{ij}. By repeatedly sampling at random the joint density in (5) and updating each of the unknown variables we can effectively estimate these variables. This is the core of the MCMC technique.

There are three types of variables that require updating: the scale parameters, the location parameters, and y_{ik}. The scale parameters are κ_η, κ_β, λ_κ , $\kappa = 1, \ldots K$. Updating the scale parameters is done by selecting a random sample of one observation from the *Gamma(a′,b′)* distribution where:

- for κ_η, $a' = a + ½I$ and $b' = b + ½\Sigma_i(\eta_i - \mu)^2$
- for κ_β , $a' = a + ½J$ and $b' = b + ½\Sigma_j\beta_j^2$
- for λ_κ , , $a' = a + ½I$ and $b' = b + ½\Sigma_i(\alpha_{ik} - \nu_\kappa)^2$

The location parameters are η_i, μ, β_j , ν_κ , and α_{ik} . These parameters are updated by the selection of a random observation from $N = (\frac{B}{A}, \frac{1}{A})$ where:

- for η_i , $A = \kappa_\eta + J$, $B = \mu\kappa_\eta + \Sigma_j(y_{ij} - \beta_j - \Sigma_k \alpha_{ik} x_{jk})$
- for μ, $A = I \kappa_\eta$, $B = \kappa_\eta \Sigma_i \eta_i$
- for β_j , $A = \kappa_\beta + I$, $B = \Sigma_i(y_{ij} - \eta_i - \Sigma_k \alpha_{ik} x_{jk})$
- for ν_κ , $A = I \lambda_\kappa$, $B = \lambda_\kappa\Sigma_i \alpha_{ik}$

- for α_{ik}, $A = \lambda_k + \sum_j x_{jk}^2$, $B = \lambda_k v_k + \sum_j x_{jk} \left(y_{ij} - \eta_i - \beta_j - \sum_{k' \neq k} \alpha_{ik'} x_{jk'}\right)$

The conditional distribution of y_{ij} given all the other unknown variables is $N(\eta_i + \beta_j + \Sigma_k \alpha_{ik'} x_{jk}, 1)$. Rejection sampling was employed to sample y. That is, consecutive values from the conditional distribution were generated until the condition $Q(y_{ij}, z_{ij}) = 1$.

In the implementation of this model y_{ik} was initially set as 1 when z_{ij} was 0 and y_{ik} was set as 0 when z_{ij} was 1. The location parameters were given the value 0 and the scale parameters were assigned 1. Ten thousand iterations were then performed to update the unknown variables and discarded so that the results of the initial step were immaterial to the results. One hundred thousand more iterations were then carried out and the results of each 100th step were preserved to compile a sample size of 1000 from the posterior distribution of the unknown variables. In order to refer to members of this sample we use the notation $^{(n)}$ so that $\alpha_{3,10}^{(45)}$ refers to the 45th observation in the sample of the variable.

The MCMC model described above and the simulated data it constructs fit the data well. To see this, let us represent the estimated value of z_{ij} on the n^{th} observation as $\hat{z}_{ij}^{(n)}$. We can calculate the estimated value of how well a given student i

knows an answer j with (6).

(6) $$\hat{y}_{ij}^{(n)} = \eta_i^{(n)} + \beta_j^{(n)} + \sum_{k=1}^{3} \alpha_{ik}^{(n)} x_{jk}$$

Here the n^{th} observation in the sample is represented as $\eta_i^{(n)}$ $\beta_i^{(n)}$ and $\alpha_{ik}^{(n)}$. Following our earlier practice, $\hat{z}_{ij}^{(n)} = 1$ if $\hat{y}_{ij}^{(n)} > 0$ and $\hat{z}_{ij}^{(n)} = 0$ if $\hat{y}_{ij}^{(n)} \leq 0$. After doing this for each n we can calculate the sample mean , $\hat{z}_{ij}$ of $\hat{z}_{ij}^{(n)}$. This is a number between 0 and 1. If we divide the interval [0,1] into L=10 equally spaced subintervals, the average $\hat{z}_{ij}^{[l]}$ of the $\hat{z}_{ij}$'s that belong to subinterval l, l = 1, …, L will define a pair (i, j). We can similarly find the mean $\bar{z}_{ij}^{[l]}$ of the observed z_{ij}'s for those (i, j) pairs. Plotting $\hat{z}_{ij}^{[l]}$ against $\bar{z}_{ij}^{[l]}$ should give a straight line if the model fits perfectly. Figure 1 gives the plot. The correlation between $\hat{z}_{ij}^{[l]}$ and $\bar{z}_{ij}^{[l]}$ is 0.9675, a good fit.

[Insert Figure 1 here]

Table 1
The Three Mophosyntactic Features Analyzed

Morpheme or Morpheme combination coded	Example in mainstream English
3rd person singular *-s*	Jill eats a lot of ice-cream.
Negation	Jack did not climb that hill.
Counterfactual Conditional *if+ed*	If Jill walked to school, she would get there earlier than Jack does.

Table 2

Linguistic feature influence on student success (α_{ik}^{M}) averaged over students, and standard deviation (scoring without allowances for AAE)

linguistic feature	Mean of α_{ik}^{M}	Standard deviation of α_{ik}^{M}
3rd Person Singular -s	-0.229196	0.061056
Counterfactual Conditional if+ed	-0.169976	0.045729
Negation	-0.067129	0.037767

Table 3

The correlation between students' overall scores and α_{ik}^{M}*, the effect of a feature on a student (scoring without allowances for AAE)*

linguistic feature	Correlation	p-value	Lower bound	Upper bound
3^{rd} Person Singular -s	0.4858209	0.000464	0.233985	0.6765637
Counterfactual Conditional if+ed	0.2663448	0.062727	-0.019246	0.5117512
Negation	0.3645716	0.010850	-0.0897320	0.5878159

Table 4

Linguistic feature influence on student success (α_{ik}^{M}) averaged over students, and standard deviation (scoring with allowances for AAE)

linguistic feature	Mean of α_{ik}^{M}	Standard deviation of α_{ik}^{M}
3rd Person Singular -s	-0.126855	0.063997
Counterfactual Conditional if+ed	-0.098812	0.039549
Negation	-0.062244	0.037988

Table 5

The correlation between students' overall scores and α_{ik}^{M} *, the effect of a feature on a student (scoring with allowances for AAE)*

linguistic feature	Correlation	p-value	Lower bound	Upper bound
3rd Person Singular -s	0.5669420	0.000027	0.337112	0.7329994
Counterfactual Conditional if+ed	0.4197904	0.002975	0.1540276	0.6289105
Negation	0.1714385	0.244000	-0.1184669	0.4344123

Table 6

Linguistic feature influence on student success (α_{ik}^{M}) averaged over students, and standard deviation (scoring with allowances for AAE)

linguistic feature	Mean of α_{ik}^{M}	Standard deviation of α_{ik}^{M}
3rd Person Singular -s	-0.064945	0.067757
Counterfactual Conditional if+ed	-0.042280	0.042411
Negation	-0.080024	0.038450

Table 7

The correlation between students' overall scores and α_{ik}^{M}*, the effect of a feature on a student (scoring with allowances for AAE)*

linguistic feature	Correlation	p-value	Lower bound	Upper bound
3rd Person Singular -s	0.5742835	0.000020	0.3467253	0.7379965
Counterfactual Conditional if+ed	0.3558352	0.013060	0.0797651	0.5812069
Negation	0.1574032	0.285300	-0.1326648	0..4226361

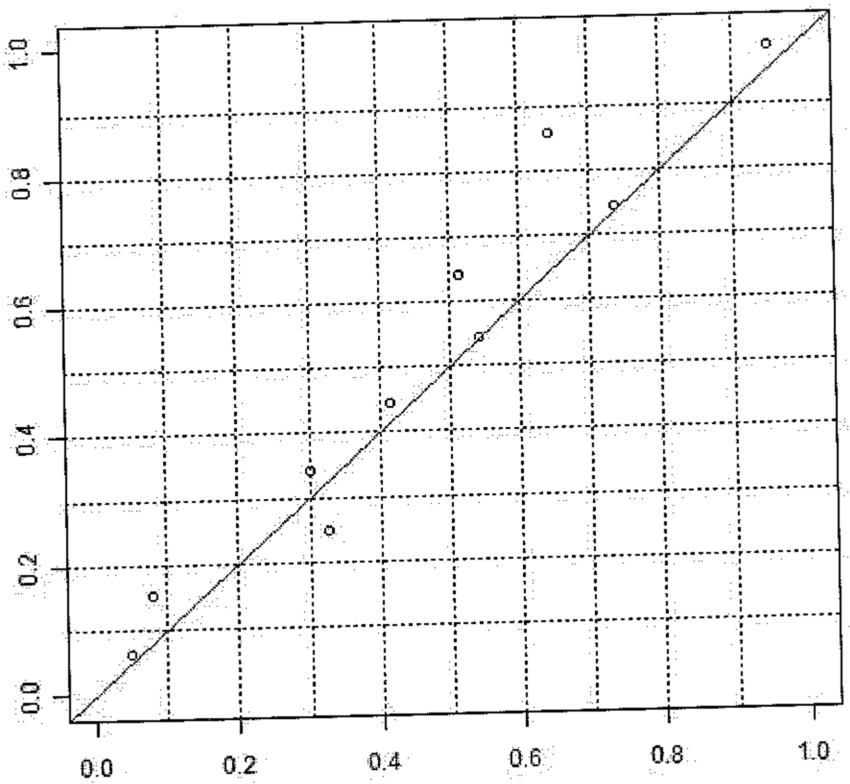

Figure 1: Goodness of Fit - Plot of $\hat{\hat{z}}^{[l]}{}_{ij}$'s against $\bar{z}^{[l]}{}_{ij}$'s